

Searching for heavy charged Higgs boson with jet substructure at the LHC

Shuo Yang^{a,b}, Qi-Shu Yan^c

^a *Physics Department, Dalian University, Dalian, 116622, P.R. China*

^b *Center for High-Energy Physics, Peking University, Beijing, 100871, P.R. China*

^c *College of Physics Sciences, Graduate University of Chinese Academy of Sciences, Beijing 100039, P.R. China*

E-mail: yangshuo@dlu.edu.cn, yanqishu@gucas.ac.cn

ABSTRACT: We study the heavy charged Higgs boson (from 800 GeV to 1500 GeV in this study) in production associated with a top quark at the LHC with the collision energy $\sqrt{s} = 14$ TeV. Such a heavy charged Higgs boson can dominantly decay into a top quark and a bottom quark due to its large Yukawa couplings, like in MSSM. To suppress background events and to confirm the signal, we reconstruct the mass bumps of the heavy charged Higgs boson and the associated top quark. For this purpose, we propose a hybrid-R reconstruction method which utilizes the top tagging technique, a jet substructure technique developed for highly boosted massive particles. By using the full hadronic mode of $pp \rightarrow H^\pm t \rightarrow tbb$ as a test field, we find that this method can greatly reduce the combinatorics in the full reconstruction and can successfully reduce background events down to a controlled level. The sensitivity of LHC to the heavy charged Higgs boson with two b taggings is studied and a 9.5σ significance can be achieved when $m_{H^\pm} = 1\text{TeV}$.

Contents

1. Introduction	1
2. Brief Review of Jet Algorithms and Jet Substructure	3
2.1 Jet Algorithms and Jet Substructure	3
2.2 Top Tagging	5
3. Heavy Charged Higgs in Full Hadronic Decay Mode	6
3.1 Heavy Charged Higgs	6
3.2 Hybrid-R Reconstruction Method for Full Hadronic Decay Mode	7
4. Analysis Results	11
4.1 Signal and Backgrounds	11
4.2 Multivariate Analysis and Search Bounds	17
5. Conclusions	19

1. Introduction

Probing the Higgs sector and unraveling the miracle of electroweak symmetry breaking (EWSB) is one of the most important goals at the Large Hadron Collider (LHC). In the Standard Model (SM), the Higgs sector includes a physical neutral Higgs boson and three would-be Goldstone bosons, while in most of extensions of the SM more than one physical Higgs bosons are predicted [1]. For example, two Higgs doublet model (THDM) extends the SM by adding one additional $SU(2)$ doublet and results in a rich phenomena in Higgs sector by boasting two CP-even Higgs bosons, h and H , one CP-odd Higgs boson A , and a charged Higgs boson H^\pm . The discovery of neutral Higgs bosons and the measurement to their properties could undoubtedly help people understand the EWSB mechanism, while the confirmation/exclusion of the existence of a charged Higgs boson at the LHC can pinpoint down the whole Higgs sector and further distinguish models.

Experiments have invested quite an effort into the search of charged Higgs bosons. There is a direct limit of $M_{H^\pm} > 78.6$ GeV from the LEP searches by its exclusive decay of $H^\pm \rightarrow \tau\nu$ and $H^\pm \rightarrow cs$ [2]. At hadron colliders, the search approaches for a charged Higgs boson differ in term of its mass range. For the low mass range $m_{H^\pm} < m_t$, the signal for a charged Higgs boson is from top quark decay $t \rightarrow H^+b$ followed by the decay $H^+ \rightarrow \bar{\tau}\nu$. At the Tevatron, due to the effects of phase space suppression for a heavy charged Higgs boson production, the search is mainly focused on the low mass range $m_{H^\pm} < m_t$. Currently there is no charged Higgs signal detected from top decay nor obvious deviation from Wtb vertex

measurement, which can put a constraint to THDM on the small and large $\tan\beta$ regions for a charged Higgs boson mass up to ~ 160 GeV [3]. However, due to its large collision energy of the LHC, the search for a heavy charged Higgs boson can be feasible via the $gb \rightarrow tH^-$ production up to a large mass range [4]. For the large mass range $m_{H^\pm} > m_t$, the signal is from the main production process the gb fusion ($gb \rightarrow tH^-$) followed by its decays. When a charged Higgs boson is heavy enough, its main decay mode can be $H^- \rightarrow t\bar{b}$ while other modes (say $H^- \rightarrow \tau\bar{\nu}$) are small and can be neglected.

When the heavy charged Higgs boson is around 1TeV or higher, the top quark from its decay can be highly boosted and the detectors of LHC may not resolve all jets from its decay when the angle separation parameter is fixed to a specific value (say $R = 0.5$ in anti-kt jet algorithm). Motivated by new technique of “jet substructure” [5, 6, 7, 8] developed for highly boosted massive particles, we propose a “hybrid-R reconstruction method” and investigate the full hadronical decay channel of the heavy charged Higgs production. Previous studies on this process focused on its semi-leptonic decay, i.e. $tH^\pm \rightarrow t\bar{b} \rightarrow jjbl\nu bb$ and mainly depend on triple- b -tagging [9, 10, 11] or four- b -tagging [12] to suppress the background. To our best knowledge, this is the first attempt to the full hadronically decaying channel for process $pp \rightarrow tH^\pm \rightarrow t\bar{b}$. Similar to its semi-leptonic decay, the full reconstruction of its full hadronic decay mode may be difficult due to its large combinatorics and the large backgrounds from $t\bar{t}$ +jets and QCD multi-jets processes [4]. Furthermore, the b jet from the boosted top may be difficult to be tagged. Therefore, we will use the top tagging and the b tagging for other isolated b jets as well as the full reconstructed objects in the final state to suppress the background.

In order to identify objects in the signal, we reconstruct all objects (i.e. two W bosons, two top quarks, and a charged Higgs boson) in the signal. In our hybrid-R reconstruction method, we first use a larger cone size for the top tagger to identify a highly boosted top. Then we utilize the fact that the b jet from the charged Higgs boson can be reliably resolved by detectors with a smaller cone size and should be the most energetic jet to reduce the combinatorics in the full reconstruction. The remaining task is to identify the associated top quark which is always produced near threshold of the process where all jets from this top quark decay can be resolved by the smaller cone. In short, our hybrid-R reconstruction utilizes the advantages of both a larger cone (for a top tagging) and a smaller cone (for other resolvable jets). And it is found that this method can work quite well when the charged Higgs boson mass is above 800GeV.

To be more precise to define our hybrid-R reconstruction method, we firstly adopt a large cone size in the top tagging method in order to capture the highly boosted top (John Hopkins top-tagger¹ [13] in Cambridge-Aachen algorithm with R optimized for each m_{H^\pm} value). Then we adopt a smaller radius R (anti- k_t with $R=0.4$) to resolve jets in the event, which is efficient to reconstruct the associated top with a small transverse momentum and the isolated b jet from the heavy charged Higgs boson decay. We introduce a few crucial cuts to suppress background and adopt the neural network and boosted decision tree to optimize cuts. It is found that, with the help of John Hopkins top-tagger and hybrid-R

¹It is a top tagging method proposed by John Hopkins group and we will introduce it in section 2.2

reconstruction method, the full-hadronic decay of a heavy charged Higgs boson at the LHC ($\sqrt{s} = 14\text{TeV}$) is reachable. We also study the sensitivity of LHC to a heavy charged Higgs boson varying from 0.8TeV to 1.5TeV and find that LHC can cover this heavy mass region when the integrated luminosity is assumed to be 100 fb^{-1} .

The remainder of this paper is organized as follows. In Sec. 2, we firstly give a brief review of jet algorithms and jet substructure. In Sec. 3, we introduce the “hybrid-R reconstruction method”. And then we employ the “hybrid-R reconstruction method” armed with top tagging method to study the process $pp \rightarrow tH^\pm \rightarrow t\bar{t}b$ with both tops in hadronic decay mode in Sec. 4. Our conclusions follow in Sec. 5.

2. Brief Review of Jet Algorithms and Jet Substructure

2.1 Jet Algorithms and Jet Substructure

A Jet algorithm is to reconstruct a jet from the collimated spray of hadrons and to make comparison between data and theoretical prediction. Jet algorithms can be casted into two categories: the cone-based algorithms and the sequential recombination algorithms. The former one relies on the intuitive idea to put a cone along the dominant direction of the energy flow. The later one repeatedly recombines the closest pair of particles which look like undoing the QCD branching. In this paper, we will focus on sequential recombination algorithms including the k_t [14, 15], the Cambridge-Aachen [16, 17] and the anti- k_t [18]. For a recent review of jet algorithms and its application at hadron collider, we refer to Ref. [5].

Sequential recombination jet algorithms begin with a list of four-momenta of pseudo-jets which can be reconstructed from the track detectors and calorimeters, both electromagnetic and hadronic ones, and recursively combine pairs of momenta into jets. The combination rules can be described as :

1. Calculate the distance d_{ij} between all pairs of pseudo-jets and the beam distance d_{iB} for each pseudo-jet.
2. Find the minimum in the list of d_{ij} and d_{iB} .
3. If the smallest entry is d_{ij} , recombine i and j into a new particle and then return to step 1.
4. Otherwise, if it is d_{iB} , define it as a final-state jet, and remove it from the original pseudo-jet list. Return to step 1.
5. Iterate this process until the original list is empty, i.e., all pseudo-jets have been clustered to jets.

The d_{ij} and d_{iB} in most popular sequential recombination algorithms for use at hadron collider can be parameterized as follows [18]:

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2, \quad (2.1)$$

$$d_{iB} = p_{ti}^{2p}, \quad (2.2)$$

where p is a parameter that is 1 for the k_t algorithm, -1 for the anti- k_t algorithm and 0 for the Cambridge-Aachen algorithm (CA algorithm). The k_t algorithm clusters soft particles first, while the anti- k_t algorithm recombines hard particles first. And the CA algorithm clusters by angle ranking. There are a few comments in order:

- There is no standard jet algorithms. The jet algorithm and the parameters could be chosen flexibly.
- All these algorithms are infrared-safe and collinear-safe.
- The anti- k_t algorithm gets circular jets area which act much like an idealized cone-based algorithm, while the k_t algorithm and CA algorithm get irregular jets area. It is one of reasons why experimental groups favor the anti- k_t from the viewpoint of detector geometry.
- The k_t and CA algorithms incline to combine the harder pseudo-jets or pseudo-jets with a large separation in the later, so it is expected that the combination in last step is the reverse of the main QCD branching or the 2-body decay of a massive particle which can be easily used in jet substructure study by iteratively tracing backwards the clustering history. However, the recombination history in anti- k_t algorithm couldn't be utilized to related the branching and to get the subjets.

The study of jet substructure has received considerable attention in recent years. At the LHC, the massive electroweak particles, i.e., W , Z , Higgs and top quarks, are always produced at a high transverse momentum ($P_t \gg M_Z$). In this case, their decay products tend to become collimated due to the large Lorentz boost factor. For highly boosted massive particles with hadronic decay, it is expected that a large cone size R could capture all the collimated decay products in a “fat jet”. And the mass of this “fat jet” is one indicator of its origin [19, 20, 21]. However, high energy QCD jets can also generate large masses. In order to suppress QCD background to an acceptable level, one have to resolve more physical information in order to pick out the “fat jet” originated from the decay of highly boosted massive particles. One way is to anatomize the substructure and to identify the subjets in a fat jet to reconstruct the delicate kinematical variables.

Along the line of this thought, the jet substructure technique has a boosting development in recent years including early studies in theory [23, 24, 25, 13, 26, 27, 28, 29, 30, 31] and in experiment [32, 33, 34, 35, 36]. For recent reviews on jet substructure, see Refs. [5, 6, 7, 8]. The earliest study of jet substructure is Ref. [23] in the context of a search for a heavy Higgs boson decaying to WW . More recently, the BDRS algorithm [25] developed a technique using the mass-drop and the filtering to transform the high- P_t $WH, ZH(H \rightarrow b\bar{b})$ channel into one of the best channels for discovery of Standard Model Higgs with small mass at the LHC. The filtering [25], i.e., reclustering the fat jet of Higgs with refined small cone size R_{filt} , is supposed to further reduce the degradation of resolution on jets caused by underlying events. Besides filtering, some new techniques including “trimming”, “pruning”, “variable R” are also developed to refine the jet resolution [29, 30, 31, 37]. Application of the BDRS method to Higgs search in the MSSM context can be found at [38, 39].

Below, we focus on the top tagging algorithms.

2.2 Top Tagging

Top quarks play an important role in understanding electroweak symmetry breaking and searching for new physics. Unlike the case at Tevatron where most of top quarks are produced near the threshold, at the LHC many boosted top quarks can be created.

In the conventional method, top quarks are reconstructed by recombining three daughters, i.e., three isolated jets in hadronic decay mode or one isolated lepton, one isolated jet and large missing energy in leptonic decay mode. And b -tagging always is used to suppress the large QCD background.

However, this conventional method confronts some problems for the top with high- P_t .

- The decay products become collimated and it is difficult to find three isolated daughters to reconstruct the top when a cone size is fixed to a specific value like $R = 0.5$ in the anti- k_t algorithm.
- The b -tagging efficiency drops significantly as the P_t increases because the tracks become too close to clearly identify the secondary vertex from the B decays in the silicon detector.
- For its leptonic decay, it is a task to determine the missing energy in the hadron collision environment.

So, the jet substructure technique used to identify boosted tops from its hadronic decay have been developed in recent years. As summarized in Ref. [5], many discriminating variables are used on subjets for top tagging, such as d_{ij} -types variables [34, 35, 40], P_t -ratio type variables [13, 26, 28, 30, 31, 40], event-shape variables [41, 42], and constraints on a W -subjet mass [13, 26] as well as other subjets correlation variables (a helicity angle θ_t [13]). In addition, the HEPTopTagger [28, 43] is designed for moderately boosted tops and some techniques are developed for leptonic top tagger [44].

Here, we adopt the top tagging method for the highly boosted top in hadronic decay proposed in Ref. [13]. We will dub it as “JH top-tagger” for simplicity and use it in later analysis. As well-known, a boosted top from its hadronic decay looks like a fat jet with three hard cores. Similar to the BDRS jet substructure method for boosted Higgs, the JH top-tagger firstly uses a large cone to cluster the event in order to capture all the decay products and relevant radiations of a boosted top and then declusters the top jet to find three subjets.

To resolve a fat top jet into the relevant hard substructure, the following recursive procedure are applied in JH top-tagger [13]:

1. Four momenta of particles is clustered with a large cone size R (CA algorithm are used in original paper)
2. The last cluster is undone to get two objects j_1 and j_2 . If the P_t ratio of the softer jet j_2 over the original jet j is too small, i.e., $P_{t_{j_2}}/P_{t_j} < \delta_p$, throw the softer j_2 and go on to decluster on the harder one.

3. The declustering step is repeated until two separated hard objects are found. If any criterion below are satisfied, the event is failed: 1) both objects are softer than δ_p (2) the two objects are too close, $\Delta\eta + \Delta\phi < \delta_r$ (3) the original jet is considered irreducible.
4. Declustering repeatedly on these two subjets will result in 2,3, or 4 hard objects.
5. The case with 3 or 4 subjets are kept. And then require the these subjets that the total mass should be near m_t , the mass of two subjets should be in the m_W window, W helicity angle θ_t should be consistent with a top decay².

The parameters involved in the method can be optimized event by event [13]. In our study, these parameters are fixed as below:

$$\delta_p = 0.19 \quad \delta_r = 0.1 \quad |\theta_t| < 0.65 . \quad (2.3)$$

An interesting application of the Higgs tagging method as well as the top quark tagging method for the heavy top partner search can be found at [45].

3. Heavy Charged Higgs in Full Hadronic Decay Mode

3.1 Heavy Charged Higgs

For a heavy charged Higgs boson ($m_{H^\pm} > m_t$), the main production process is the gb fusion process $gb \rightarrow tH^-$. And the LHC can probe the charged Higgs boson in a wide mass range up to TeV scale. Additional productions including $W^\pm H^\mp$ [47], ΦH^\pm ($\Phi = A, h, H$) [48] and $H^+ H^-$ [49] have also been studied but the cross sections are generally small in these modes. For the H^\pm with $m_{H^\pm} > m_t + m_b$, the decay $H^- \rightarrow \bar{t}b$ is the main decay mode due to the large Yukawa coupling. Although the decay mode $H^- \rightarrow \tau\bar{\nu}$ is an interesting one, its small branching ratio makes it difficult to probe a heavy charged Higgs boson. The early studies on the search for a charged Higgs on $pp \rightarrow tH^\pm$ production can be found in Ref [9, 10, 11, 12]. The search in this channel are also carried out by both ATLAS and CMS collaborations. These collaborations mainly investigate $gb \rightarrow tH^- \rightarrow t\bar{t}b \rightarrow bq\tau(had)\nu bb$ and the semi-leptonical channel $gb \rightarrow tH^- \rightarrow t\bar{t}b \rightarrow bqql\nu b$. For search in τ channel $gb \rightarrow tH^- \rightarrow t\bar{t}b \rightarrow bq\tau(had)\nu b$, the branching ration of $H^\pm \rightarrow \tau\nu$ is small especially for higher H^\pm mass. The main backgrounds to this signal are $t\bar{t}$ events, W +jets and QCD multi-jet events [4]. For semi-leptonically decaying channel of H^\pm production, i.e., $tH^\pm \rightarrow t\bar{t}b \rightarrow jjbl\nu bb$, the main backgrounds come from $t\bar{t}$ + jets, $t\bar{t} + b\bar{b}(QCD)$, and $t\bar{t} + b\bar{b}(EW)$ [4]. These studies indicate the search for a heavy charged Higgs boson from this channel has to rely on triple- b -tagging [9, 10, 11] or four- b -tagging [12] in order to suppress the background.

²The helicity angle θ_t is defined as the angle, measured in the rest frame of the reconstructed W , between the reconstructed top's flight direction and one of the W decay products. The lower P_t subjet in the lab frame are chosen to set the angle.

Armed with the jet substructure method to tag highly boosted top quarks, we investigate the full hadronic decay mode for process $pp \rightarrow gb \rightarrow t_2 H^- \rightarrow t_2 \bar{t}_1 b_3$ ³. To our best knowledge, this is the first attempt for such a study in literature. For convenience, we here call the top quark and the b quark from the charged Higgs boson decay as t_1 and b_3 , respectively, and the associated top quark as t_2 as well. We attempt to reconstruct all objects in the final states. For this purpose, we resort to the top tagging method and a “hybrid-R reconstruction method” which can help to overcome the combinatorics problem in the full reconstruction while maintaining a remarkable suppression to backgrounds (especially QCD background).

Since we only concentrate on the study of the interaction of the heavy charged Higgs with top and bottom quarks, we simplify the Lagrangian as follows:

$$\begin{aligned} \mathcal{L} = & \mathcal{L}_{SM} + \partial_\mu H^\pm \partial^\mu H^\mp + m_{H^\pm}^2 H^\pm H^\mp \\ & + H^\pm \bar{t}(Y_L P_L + Y_R P_R)b + h.c. \end{aligned} \quad (3.1)$$

The couplings Y_L and Y_R are treated as free parameters in the model-independent analysis. In two Higgs doublet model, for simplicity, we assume the branching ratio of charged Higgs into tb is 100% and take $Y_L = Y_R = 1$.

3.2 Hybrid-R Reconstruction Method for Full Hadronic Decay Mode

For the full hadronically decaying of process $pp \rightarrow gb \rightarrow t_2 H^- \rightarrow t_2 \bar{t}_1 b_3$, the main backgrounds are $t\bar{t}$ and QCD multi-jets events. We focus on a heavy charged Higgs boson with mass around 1 TeV or so.

In such a mass region, the daughter t_1 from the heavy charged Higgs boson decay is always highly boosted. And a larger cone size R is expected to capture all the hadronic decay products of t_1 in a fat jet. However, the associated t_2 in the $pp \rightarrow t_2 H^\pm$ process is produced near its threshold and its decay products generally can not form a fat jet. So the large cone size used for t_1 tagging fails to catch the t_2 and only a small cone size in a jet algorithm can resolve all its three jets. Then the traditional top reconstruction method by recombining three jets is expected to be efficient to tag t_2 .

In order to reconstruct all objects in the process $pp \rightarrow gb \rightarrow t_2 H^- \rightarrow t_2 \bar{t}_1 b_3$ and to exploit the advantages of a larger and a smaller cone size jet algorithms, we propose a “hybrid-R reconstruction method”.

The five steps we used in our “hybrid-R reconstruction method” can be described as follows:

1. At the preselection level, we adopt the CA jet algorithm with a larger cone size parameter and use a top tagger method to capture a highly boosted top and preselect events. The JH top-tagger is employed in this work. The jets at this step form the jet set J_0 .
2. After capturing a top jet labeled as t_1 , we re-cluster the pseudo-jets in the event with a smaller cone size parameter (the anti- k_t algorithm with $R = 0.4$ is chosen as an example) and then get a new jet list L_0 .

³We also include $pp \rightarrow gb \rightarrow \bar{t}_2 H^+ \rightarrow \bar{t}_2 t_1 \bar{b}_3$ in our calculation

3. If a small-size jet is within the larger cone of the direction of the tagged top jet, we remove it from the jets list for the b_3 and t_2 reconstruction. The rest of jets form a new jet list L_1 .
4. We identify the most energetic jet in the list L_1 as a b_3 jet. This choice also implies that the tagged top t_1 and the identified b_3 fly back-to-back. This is due to the observation that the tagged top in the larger cone size by the top-tagger method contains subjets which are massive and energetic and the leading two objects with large transverse momentum in the smaller cone size always fly back to back. So we don't impose a cut on $\Delta\phi$ or ΔR here. The unused jets form the list L_2 .
5. We require there are at least 3 jets in the jet list L_2 in an event. Then we introduce a χ^2 method to tag the second top quark in the signal. Here, $\chi^2 = \frac{|m(j_1, j_2) - m_W^{\text{PDG}}|^2}{\sigma_W^2} + \frac{|m(j_1, j_2, j_3) - m_t^{\text{PDG}}|^2}{\sigma_t^2}$ where σ_W and σ_t are equal to 10 GeV and 15 GeV, respectively, which is to account for the detector resolution capability. The m_W^{PDG} and m_t^{PDG} are taken as 80.5 GeV and 173 GeV, respectively. The combination which minimize the χ^2 is chosen as the right reconstruction.

In step 2 we recluster a selected event with a smaller cone size, and such a reclustering has also been used to reduce the degradation of the resolution on a fat jet from highly boosted massive particle in filtering [25] and some other grooming techniques, and also can be used as preselection rule [28], and can be used to separate top jets from QCD jets in the hollow cone method [46] as well. But it should be stressed that, in our hybrid-R reconstruction method, the reclustering and removing the smaller jets within the radius of the tagged top is to catch the highly boosted top and to reconstruct successfully the other massive particle with a smaller P_t (Here it is the t_2).

Another remarkable point is that any top tagging algorithm for highly boosted top can be used in the first step although the JH top-tagger is used as an example.

Before carrying on the detailed analysis, we would like to show some features of the signal which support our hybrid-R reconstruction method. At parton level, in Fig. 1, we show the distribution of transverse momenta of t_2 , t_1 , and b_3 , and the angle separation among two of t_1 , t_2 and b_3 in the case of $M_{H^\pm} = 1$ TeV. The distributions of $P_t(t_1)$ and $P_t(b_3)$ are similar and yield an important clue for $P_t(t_1)$ and $P_t(b_3)$. We find that $P_t(t_1) = P_t(b_3) \geq \frac{3}{10}m_{H^\pm}$ can work and so we will use it in our simple cut method which will be introduced in the following section. We also show the distribution of $P_t(t_1) - P_t(t_2)$ and $R(t_1 b_3) - R(t_2 b_3)$, which provide important hints to determine the right combination. As shown in Fig. 1, one can find the fact that in most of phase space the transverse momentum of t_1 is larger than t_2 . Meanwhile, the angle separation between t_1 and b_3 demonstrates that these two objects indeed fly back-to-back. And in most of cases, the angle separation between t_1 and b_3 is larger than that between t_2 and b_3 . At parton level, it is also found that the $\Delta\phi(t_1, b_3)$ is always close to π , which may be useful in the full reconstruction.

At the fast detector simulation level by using PGS4 [50], as demonstrated in the left panel of Fig. 2, we show the mass distribution of the leading massive jet, which is assumed

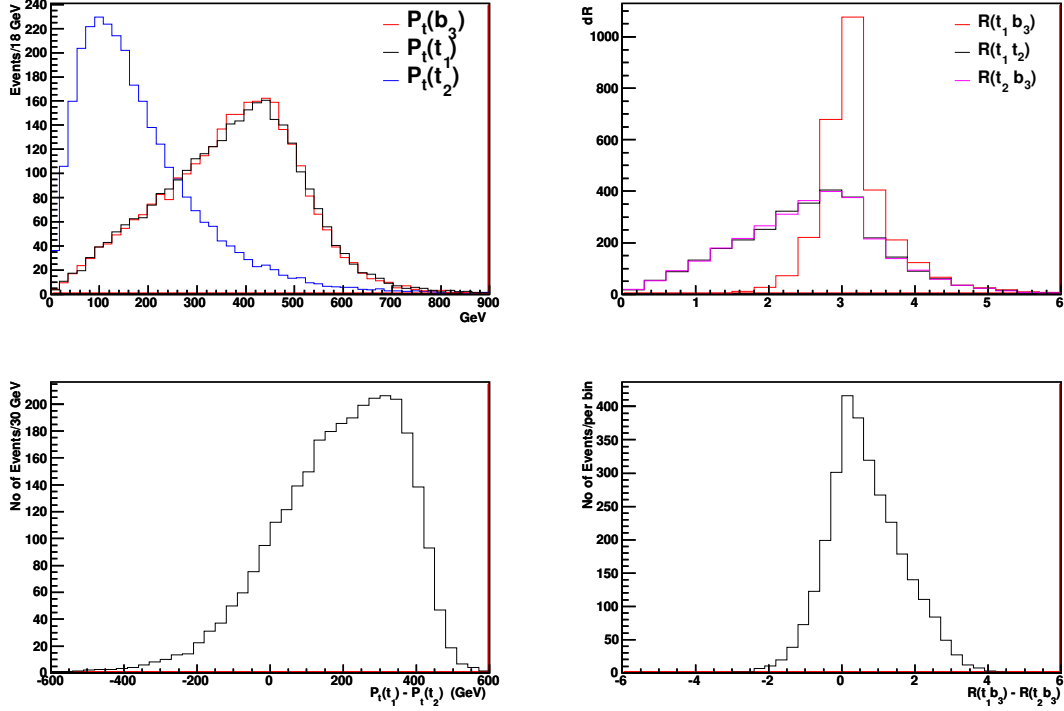


Figure 1: The transverse momenta and the angle separation are exposed at parton level with $m_{H^\pm} = 1.0\text{TeV}$. The transverse momentum difference between t_1 and t_2 and the angle separation difference between $R(t_1 b_3)$ and $R(t_2 b_3)$ are also shown.

to be a t_1 jet in our hybrid-R reconstruction method. Then from the rest of jets, we pick a jet with the largest transverse momentum as the b_3 jet. After picking out these two jets, we examine the invariant mass of the combined Lorentz vector from them, which is also displayed in the right side of Fig. 2. We examine how these two types of mass distribution vary to the change of angle separation parameter R in the anti- k_t algorithm.

In Fig. 2, we can read out that when $R = 0.4$, neither a W boson nor a top quark can be captured in a single massive jet significantly. When $R = 0.7$, a considerable fraction of W boson can be captured in one single massive jet while a small fraction of t quark can be captured. When $R = 1.0$ and $R = 1.3$, quite a fraction of top quark can be captured by a single massive jet.

Meanwhile, when R is large enough, the simple reconstruction method by using the most-massive jet and the most-energetic jet can even reconstruct the heavy charged Higgs boson mass bump, as demonstrated in Fig. 2. Both parton level and fast detector simulation level analysis demonstrate that our event reconstruction method captures the most important feature of the signal.

In our Monte Carlo study, the signal events are generated by Madgraph/MadEvent [51] and background events are produced by Alpgen [52]. These events are fed to DECAY to generate full hadronic decay final states and pass to PYTHIA [53] to simulate showering,

fragmentation, initial state radiation, final state radiation, and multi-interaction as well. After that, we use fastjet [54] and SpartyJet[55] to perform jet clustering and analysis.

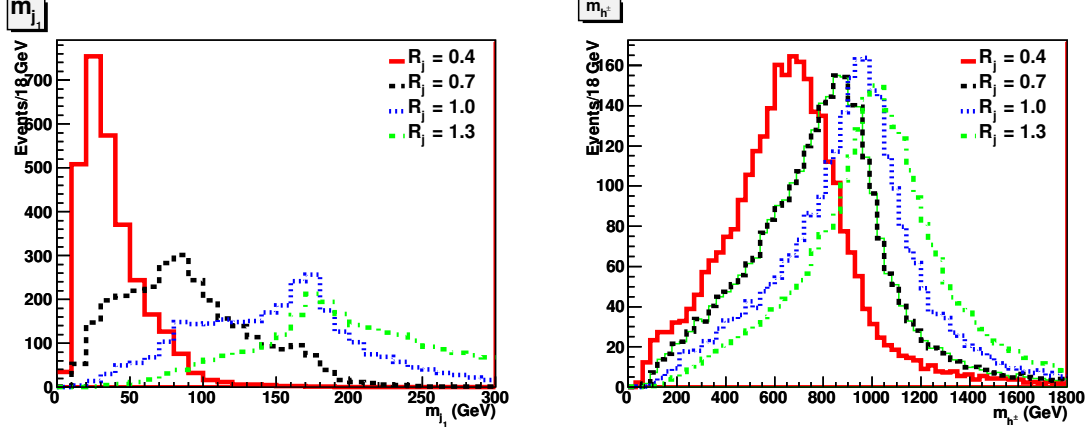


Figure 2: The mass distribution of the leading massive jet and the reconstructed charged Higgs boson mass are shown with the case $m_{H^\pm} = 1\text{TeV}$. The dependence on the angle separation parameter R in the anti- k_t algorithm is also exposed. To be more realistic, the detector effects have been taken into account by using PGS4.

Now, we examine the top tagging efficiency of JH top-tagger for different values of the cone size parameter R in both k_t and CA algorithms by using the SpartyJet package. The results are tabulated in Table 1 and Table 2, which are two cases study for $m_{H^\pm} = 1.0\text{TeV}$ and $m_{H^\pm} = 1.5\text{TeV}$, respectively. As shown in Table 1 and Table 2, the tagging efficiency of the CA algorithm is better than that of the k_t algorithm in a larger R , especially for a larger value of m_{H^\pm} . Another remarkable point is that the efficiency of the CA algorithm does not depend upon the value of R as that of the k_t algorithm does in large R range. Due to these feature, in later analysis, we will use the top-tagger in CA algorithm with optimized cone sizes for different m_{H^\pm} .

R	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
k_t algorithm	16%	22%	25%	28%	29%	29%	28%	27%
CA algorithm	13%	19%	24%	27%	29%	30%	30%	30%

Table 1: The efficiency of top tagger in k_t and CA algorithms along with the variation of angle separation parameter R are tabulated. The charged Higgs boson mass is assumed to be 1.0 TeV.

This behavior can be understood by the clustering sequence of jet algorithms. In the CA algorithm, the clustering is performed by angle ranking and the softer drop method, i.e., δ_p cut in the top-tagger will efficiently remove the softer branching, underlying events and pile-up with a large separation to the hard subjects. Therefore it is successful to find the refined radius for each hard core in a top jet. However, in the k_t algorithm, the clustering recombines the hardest and largest separated particles in the last stage of recombination, the softer pseudojets have been merged ahead and the δ_p cut actually don't work as it

R	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4
k_t algorithm	25%	27%	28%	28%	27%	26%	24%	22%
CA algorithm	23%	27%	29%	30%	30%	30%	30%	29%

Table 2: The efficiency of top tagger in k_t and CA algorithms along with the variation of angle separation parameter R are tabulated. The charged Higgs boson mass is assumed to be 1.5 TeV.

does in the CA algorithm. Additionally, in the k_t algorithm, a jet has a larger area and a stronger P_t dependence than that in the CA algorithm [56].

These features also explain why the efficiency of top-tagger in the CA algorithm with a smaller R is smaller than that in the k_t algorithm. In a smaller cone size, the CA algorithm can not capture all jets from the top quark decay while the k_t algorithm is better since it can bring more protojets and get a larger mass for its larger jet area and no softer drop.

It is also remarkable that the configuration of the b quark and the W boson from t_1 decay flying near to the direction perpendicular to the boost direction of t_1 has larger probability to form a top jet. While the configuration either the b or the W boson flying parallel to the boost direction can not form a top jet. The parameters of top tagging given in Eq. (2.3) also affect the tagging efficiency. This explains why there are more than 70% or so in the signal can not be successfully tagged as a top jet.

4. Analysis Results

In this section, we present our detailed analysis. One obvious question is whether signals can be recorded at detectors of LHC. Below we use the triggering system of ATLAS detector as an example to examine whether the signals can be recorded.

At the level-1 triggering, signals can be selected out by the jet triggering from heavy charged Higgs boson decay which have large P_t . The multi-jet trigger should also help since there are more than 3 jets in the signals. Furthermore, the scalar sum of transverse energy, which should be close to the heavy charged Higgs boson mass, can also help to pick out the signals. At the level-2 triggering, the tagged b jets can be helpful to record signals. At the event filter level, the top tagger algorithm can be implemented to select signals.

At the offline analysis level, we introduce a set of cuts and compare their efficiencies between the signal and backgrounds. Multivariate analysis methods, the neural network and the boosted decision tree, are also used to improve the significance.

4.1 Signal and Backgrounds

The cross section of the process $pp \rightarrow gb \rightarrow t_2 H^- \rightarrow t_2 \bar{t}_1 b_3$ and the decay width of heavy charged Higgs boson are shown in Fig. 3. The cross section is about 0.2pb for $m_{H^\pm} = 0.8\text{TeV}$, which drops down to 0.02pb for $m_{H^\pm} = 1.5\text{TeV}$. While the decay width is about 75GeV for $m_{H^\pm} = 0.8\text{TeV}$, which increases to 170GeV for $m_{H^\pm} = 1.5\text{TeV}$. It is found that the decay width can be approximated by $\Gamma_{m_{H^\pm}} \approx 10\% \times m_{h^\pm}$. After using the PGS4 to simulate the detector effect, it is also found that the effect of decay width to the mass bump of H^\pm is roughly comparable to the detector effect, which means the mass

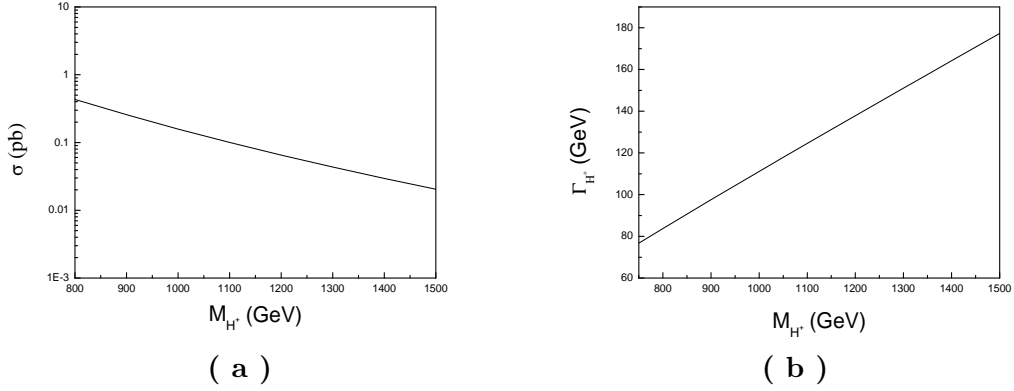


Figure 3: (a) The cross section as a function of M_{H^\pm} at the LHC with $\sqrt{s} = 14$ TeV is plotted. (b) The width as a function of M_{H^\pm} is shown.

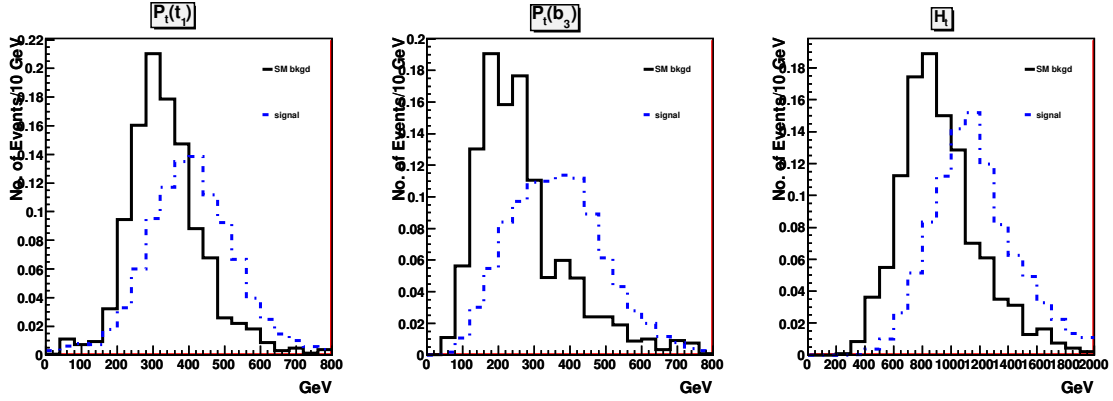


Figure 4: For the case $m_{H^\pm} = 1$ TeV, the normalized distributions of signal and backgrounds for kinematic observables, $P_t(t_1)$, $P_t(b_3)$, and H_t , are demonstrated, respectively.

bump of the charged Higgs boson is not too broad to be detectable from the detectors of the LHC.

Since the cross section of SM background processes is huge, it is necessary to use the b -tagging (assuming b -tagging efficiency 60% and mis-tagging efficiency for g and quarks 1%) to suppress the SM background events. We demand that the most energetic isolated jet and one extra b jet in the jet list L_0 must be tagged as b jets. The most energetic isolated jet must be a tagged b jet can suppress the $t\bar{t}$ + jets background by a factor of 100 since the isolated jet in this process is mainly from gluon radiation.

Below we introduce a few crucial kinematic observables. One of the most characteristic features of the signal is that the t_1 and b_3 from the decay of heavy charged Higgs boson are very energetic and fly in the small η region. Therefore, the top-tagger is efficient to capture the highly boosted t_1 coming from the decay of the heavy charged Higgs since the large mass of m_{H^\pm} will transform to the high P_t of the top quark t_1 . For $t\bar{t}j$ background, there is only a fraction of events could provide a high P_t top jet passing the JH top-tagger.

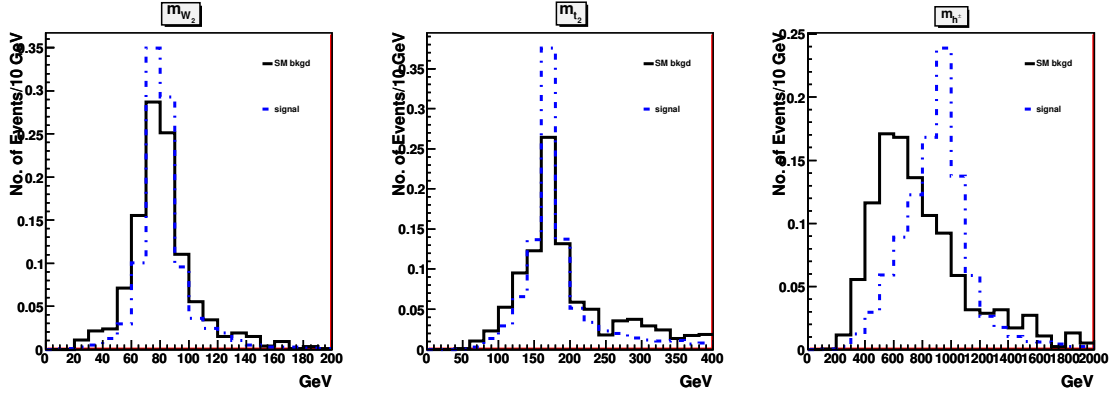


Figure 5: For the case $m_{H^\pm} = 1\text{TeV}$, the normalized distributions of both signal and background for reconstructed mass peaks, m_{W^2} , m_{t^2} , and m_{H^\pm} are shown.

However, it is difficult to find a massive jet has the same substructure as the highly boosted top in QCD multi-jet events. Meanwhile, the H_T observable, i.e, the scalar P_t sum of final states, should be a good discriminant for signal and background separation.

After tagging a highly boosted t_1 , for the signal the b_3 jet generally should be the most energetic jet in the rest of jet list L_1 . It is also expected that the distribution of transverse momentum of t_1 and b_3 should be good for signal and background separation. Once both t_1 and b_3 are identified, the mass bump of the heavy charged Higgs boson can be reconstructed. Then the only combinatorics is from the reconstruction of the second top quark, t_2 , which can be reconstructed from the remain jet list L_2 . While it is difficult for backgrounds to reconstruct such a top quark especially for QCD multi-jet. So it is obvious that the reconstruction mass bumps from t_2 including m_{W^2} and m_{t^2} are important to separate the signal and backgrounds.

In Figs. 4-5, distributions of kinematic variables of events which pass the preselection rules are presented. As shown in Fig. 4, the distribution of $P_t(t_1)$, $P_t(b_3)$, and H_T for both signal and background events are shown. In Fig. 5, the mass bump of m_{W^2} , m_{t^2} , and m_{H^\pm} are shown.

According to these kinematic features, a set of preselection cuts are chosen and the comparison of the cuts efficiencies between the signal and backgrounds are shown in Table 3 and Table 4. These two tables correspond to the case for heavy charged Higgs with mass 1 TeV and 1.5 TeV at the LHC ($\sqrt{s}=14\text{ TeV}$), respectively. Due to the preselection rules introduced in our reconstruction method, the more relevant QCD multijet background are generated by requiring the leading four jets with $P_{t,min} > 100\text{GeV}$.

Some of the preselection rules, which have been encoded in our reconstruction method, are listed as

- Detector acceptance cuts are chosen as $P_t(j) > 20\text{ GeV}$, $|\eta|(j) < 2.5$.
- A cut on the number of jets in the jet list L_0 , $n_j \geq 6$.
- A cut on the H_T , $H_T > \frac{7}{10}M_{H^\pm}$, and the centrality, $C > 0.3$.

- A cut on the leading energetic jet in the jet list L_0 , $E(j_1) > \frac{7}{20}M_{H^\pm}$.
- A cut on the second leading energetic jet in the jet list L_0 , $E(j_1) > \frac{1}{4}M_{H^\pm}$.

At the data analysis level, for the sake of comparison with the multivariate analysis methods, we introduce a simple cut method consisting of several important cuts after the full reconstruction, which are listed below

- A large P_t requirement on t_1 jet captured by the top-tagger and on the supposed b_3 in hybrid-R reconstruction method, i.e. $P_t(b_3) > \frac{3}{10}M_{H^\pm}$ and $P_t(t_1) > \frac{3}{10}M_{H^\pm}$.
- Cuts are chosen for the mass bumps in the t_2 reconstruction, i.e., $|m_{W_2} - m_W^{\text{PDG}}| < 20$ GeV and $|m_{t_2} - m_t^{\text{PDG}}| < 30\text{GeV}$.
- The charged Higgs boson mass window is chosen as $|m_{H^\pm} - m_{H^\pm}^{\text{assumed}}| < 200$ GeV.

The efficiency of each cut is also shown in Tables 3 and 4. The effects of b taggings, which can help to reduce combinatorics and suppress QCD background events, are also shown. Here the b tagging requires the leading P_t jet must be tagged, where we assume b -tagging efficiency as 0.6 and the mistagging efficiency as 0.01.

	signal $m_{H^\pm} = 1$ TeV	$t\bar{t}$ + jets	QCD $n_j \geq 4$
Cross Section x Br (pb)	0.053	553	9186.8
After JH tagger & $H_T > 400\text{GeV}$	29%	7%	4.8×10^{-3}
Two b -taggings	10%	4.2×10^{-4}	4.8×10^{-7}
The number of jets in jet list L_0 , $n_j \geq 6$	10%	3.7×10^{-4}	3.4×10^{-7}
$H_T > 700\text{GeV}$ & $C > 0.3$	9.6%	2.1×10^{-4}	3.0×10^{-7}
The leading jet $E(j_1) > 350\text{GeV}$	8.9%	1.6×10^{-4}	2.5×10^{-7}
The second leading jet $E(j_2) > 250\text{GeV}$	7.9%	1.3×10^{-4}	2.2×10^{-7}
$P_t(b_3) > 300$ GeV & $P_t(t_1) > 300$ GeV	5.3%	4.5×10^{-5}	7.7×10^{-8}
$ m_{W_2} - m_W^{\text{PDG}} < 20$ GeV & $ m_{t_2} - m_t^{\text{PDG}} < 30$ GeV	3.1%	2.6×10^{-5}	2.3×10^{-8}
$ m_{H^\pm} - m_{H^\pm}^{\text{assumed}} < 200$ GeV	2.5%	1.5×10^{-5}	1.5×10^{-8}
Events in 100 fb^{-1}	133	830	13.8

Table 3: The efficiencies of both signal and background before and after using JH top tagger are displayed. The collision energy is assumed to be 14 TeV. The JH top tagger and CA algorithm are used with the angle separation parameter $R = 1.1$.

As shown in Table 3, the top tagging efficiency is 30% for signal while the mistagging efficiency for QCD background are at 0.1% level. For $t\bar{t}j$ background, there is a small fraction about 7% where one of the top quark can be boosted. As for $M_{H^\pm} = 1.5$ TeV, the top tagging efficiency and the mistagging efficiency are similar. The comparison of the cuts efficiencies for the signal and backgrounds are shown step by step. After imposing

	signal $m_{H^\pm} = 1.5 \text{ TeV}$	$t\bar{t} + \text{jets}$	QCD $n_j \geq 4$
Cross Section x Br (pb)	0.007	553	9186.8
After JH tagger & $H_T > 400\text{GeV}$	29%	3%	3.3×10^{-3}
Two b -taggings	10%	1.8×10^{-4}	3.3×10^{-7}
The number of jets in jet list L_0 , $n_j \geq 6$	9.3%	1.7×10^{-4}	2.5×10^{-7}
$H_T > 1050\text{GeV}$ & $C > 0.3$	8.3%	5.5×10^{-5}	8.9×10^{-8}
The leading jet $E(j_1) > 525\text{GeV}$	7.7%	4.4×10^{-5}	7.3×10^{-8}
The second leading jet $E(j_2) > 375\text{GeV}$	6.8%	3.4×10^{-5}	5.4×10^{-8}
$P_t(b_3) > 450 \text{ GeV}$ & $P_t(t_1) > 450 \text{ GeV}$	4.4%	9.2×10^{-6}	8.5×10^{-9}
$ m_{W_2} - m_W^{\text{PDG}} < 20 \text{ GeV}$ $ m_{t_2} - m_t^{\text{PDG}} < 30 \text{ GeV}$	2.5%	4.6×10^{-6}	4.3×10^{-9}
$ m_{H^\pm} - m_{H^\pm}^{\text{assumed}} < 200 \text{ GeV}$	1.8%	1.9×10^{-6}	1.8×10^{-10}
Events in 100 fb^{-1}	13	105	0.2

Table 4: The number of events of both signal and background before and after using JH top tagger are displayed. The collision energy is assumed to be 14 TeV. The JH top tagger and CA algorithm are used with the angle separation parameter $R = 0.9$.

the requirement of successful reconstruction of t_2 , the background have been considerably rejected. Furthermore, if a mass window cut on the reconstructed charged Higgs boson is imposed, the background is significantly suppressed by a factor 3 or so for the SM background while the signal is retained in 50% (for $M_{H^\pm} = 1\text{TeV}$). It is of interest to note that the cuts used in above steps can suppress the background by a factor 10 (100) for $m_{H^\pm} = 1\text{TeV}$ ($m_{H^\pm} = 1.5\text{TeV}$) at least. This suppression factor increases with the increase of charged Higgs boson mass.

We would like to make a comment from Table 3 and Table 4. These two Tables provide crucial evidences to use the top tagging substructure technique and b taggings. Without the top tagger and b taggings to select the most relevant events, the computing and analysis time for background events will increase drastically, which might be difficult to make the full hadronic decay mode realistic. Furthermore, if the standard top reconstruction methods are used for t_1 reconstruction, we have to confront with the combinatorics problem. Even the most energetic top quark t_1 can be successfully reconstructed, it is found that the dominant background $t\bar{t} + \text{jets}$ can not be sufficiently suppressed. Similarly, without the top tagging, the mistagging rate from QCD background can reach to a few percent. For the sake of comparison, we have considered a full reconstruction without using the top tagger. We have considered events with number of jets $n_j \geq 7$ and have tried a biased χ^2 method, where χ^2 is defined as

$$\chi^2 = \frac{|m_{W_1} - m_W^{\text{PDG}}|^2}{\sigma_W^2} + \frac{|m_{W_2} - m_W^{\text{PDG}}|^2}{\sigma_W^2} + \frac{|m_{t_1} - m_t^{\text{PDG}}|^2}{\sigma_t^2} + \frac{|m_{t_2} - m_t^{\text{PDG}}|^2}{\sigma_t^2} + \frac{|\delta\phi(t_1, b_3) - \pi|^2}{\sigma_\phi^2}. \quad (4.1)$$

In order to reduce combinatorics and to save processing time, among the leading two jets of each event, we assume one must be b_3 and the other must from t_1 decay. The last term in Eq. (4.1) is to capture the configuration with t_1 and b_3 from the heavy charged Higgs boson decay, and we have used $\sigma_\phi = 0.2$. Then the mass of the heavy charged Higgs boson can be reconstructed. Without the help of TMVA methods, for $m_{H^\pm} = 1$ TeV ($m_{H^\pm} = 1.5$ TeV) case, we find the significance can not be better than 4 (1.5) due to a large number of background events and the deterioration of wrong reconstruction of signal events. While when top tagger is used, we can greatly reduce combinatorics and analysis time and maintaining a better significance. Therefore the top tagging methods adopted here are indeed necessary. It is also remarkable that the b taggings are also crucial to suppress the SM background (especially the QCD background) to a computable and controllable level. Without b taggings, the significance can not be better than 1.2(0.5).

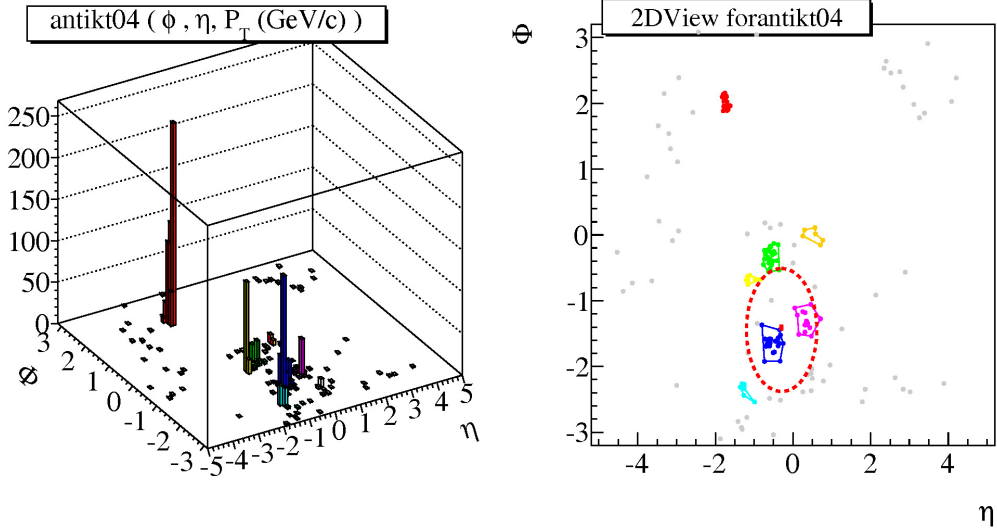


Figure 6: The lego plot of P_t and 2-Dimension plot of $\eta - \Phi$ for a typical signal event clustering with anti- k_t $R = 0.4$ for illustration. And the direction of the top jet tagged by JH top-tagger and the top area with $R = 1.0$ are marked with a red circle.

In Fig. 6, we also show the P_t lego plot and 2-Dimension $\eta - \phi$ diagram of a typical signal event (anti- k_t $R=0.4$) passing our selection rules. The direction of the top jet tagged by JH top-tagger and the top area with $R = 1.0$ are marked with a red circle. In this

example, the traditional recombination method fails because only two jets can be found with the jet algorithm parameter $R=0.4$ within the tagged top jet. However, such a signal event can be efficiently selected out by our hybrid-R reconstruction method equipped with the top tagging algorithm. Furthermore, our method is efficient to reconstruct the other top because we reduce substantially the combinatorics after subtracting the jets within the tagged top and removing the b_3 jet. For a naive comparison, the combinatorial number in the traditional recombination method is $C_7^3 = 35$ to reconstruct a top in one event with seven jets. In our method, the combinatorial number is $C_4^3 = 4$.

4.2 Multivariate Analysis and Search Bounds

In this subsection, we explore the multivariate analysis (TMVA) to optimize the cuts and improve the sensitivity. Both of the neural network discriminant analysis and the boosted decision tree discriminant analysis are considered. TMVA methods have been broadly used in the top quark precision measurements by experimental collaborations [57, 58, 59] and W-jet substructure analysis [60], as well as new physics search studies, like Higgs boson search [61] and heavy quarks search [62, 63, 64]. The boost decision tree method has been adopted in the light charged Higgs boson search [65].

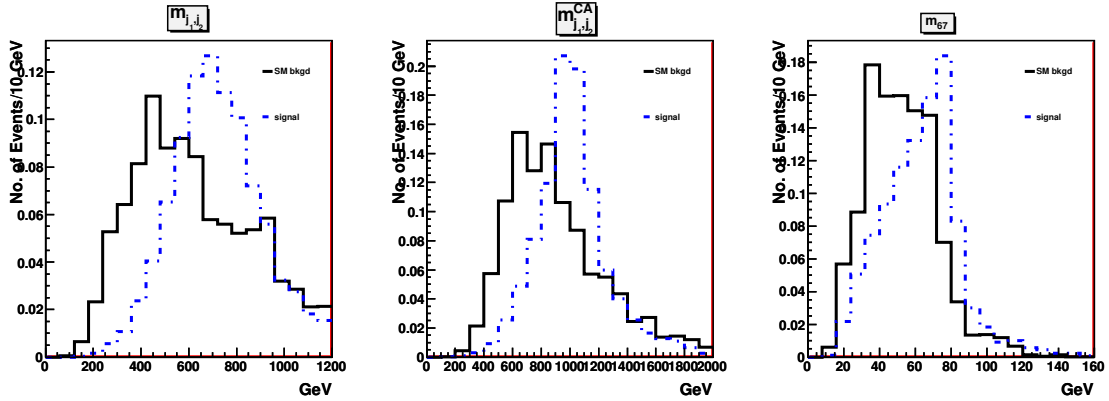


Figure 7: For the case $m_{H^\pm} = 1\text{TeV}$, the normalized distributions of both signal and background for a few kinematic observables, $m_{j_1 j_2}$, $m_{j_1 j_2}^{CA}$, and m_{67} are shown.

In Fig. (7), three extra kinematic observables are shown. Except the kinematic variables introduced in Section 4.1, we also find several extra kinematic observables helpful to separate signal and background.

- The observable m_{j_1, j_2} is the maximal invariant mass of a pair of jets in the jet list L_0 .
- The observable $m_{j_1 j_2}^{CA}$ is the invariant mass of the leading two jets in the jet list J_0 . This observable can form a bump for signals near the mass of heavy charged Higgs boson.
- The observable m_{67} is the minimal invariant mass of a pair of jets for the leading 7 jets in the jet list L_0 . The signals tend to condense near the mass of W boson.

- The maximal value of invariant mass of leading three jets in the jet list L_0 , which is labeled as m_{123}^x , is found to be concentrated near the mass of heavy charged Higgs boson for signal.
- The minimal value of invariant mass of three jets among the leading 7 jets in the jet list L_0 , which is labeled as m_{567} , is found to be concentrated near the mass of top quark for signal.
- The invariant mass of the leading 7 jets in the jet list L_0 is found to be useful for signal and background discrimination.

The discrimination distributions of multivariate analysis are demonstrated in Fig. 8. As it is shown, multivariate analysis methods (Neural Network and Boosted Decision Tree) using the input kinematic observables indeed work well to separate the signal from background events. And the efficiencies of these two TMVA methods are also listed in the Tables 5 and 6, as a comparison to that of the simple cut method used above. As shown in the Tables 5 and 6, both of the neural network and boosted decision tree methods can improve the significance by a factor 100% or more. For instance, for the charged Higgs with mass 1 TeV, both TMVA methods can improve the significance by gaining more signals at the price for a little bit more background events. This improvement is also demonstrated in the mass window of the charged Higgs boson, as demonstrated in Fig. 9.

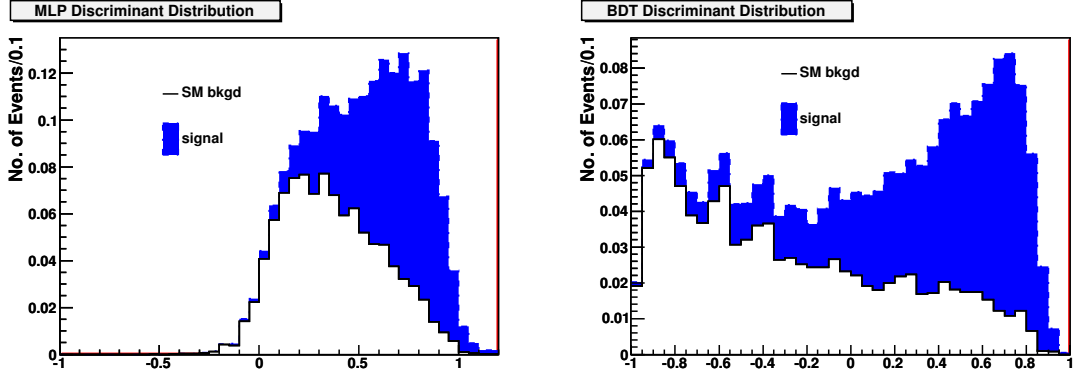


Figure 8: The discriminant distributions of neural network analysis and boosted decision tree analysis are demonstrated, where both signal and background is normalized to one.

Sensitivity of LHC to heavy charged Higgs boson in the $m_{H^\pm}-\sigma(pp \rightarrow H^\pm t)$ plane are investigated by scanning the charged Higgs boson mass m_{H^\pm} from 800 GeV to 1.5 TeV with a step 100 GeV. For each Higgs boson mass, the cone parameter R in the JH-tagger method with the best selection efficiency for signal is taken. The sensitivity reachable at LHC with 14 TeV collision energy with two b taggings is provided in Table 7 and is displayed in Fig. (10). Roughly speaking, the TMVA can improve the sensitivity by a factor 100%. In Fig. (10), we show bounds from both 100 fb^{-1} and 500 fb^{-1} integrated luminosity. Two sample points of MSSM (marked with a solid star and a solid circle, respectively) are also shown.

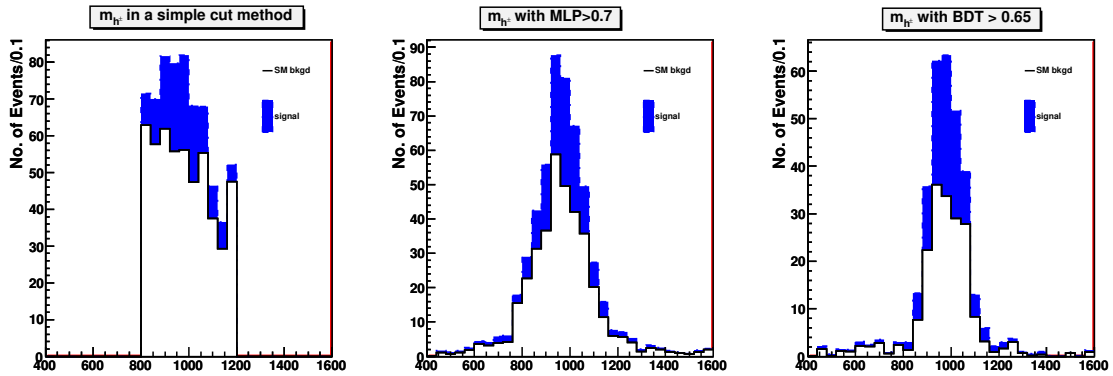


Figure 9: The reconstructed mass bump of charged Higgs boson in three analysis methods are demonstrated, where both signal and background have been normalized to one before analysis.

	signal $m_{H^\pm} = 1.0 \text{ TeV}$	$t\bar{t} + \text{jets}$	QCD $n_j \geq 4$
After simple cuts	2.5%	1.5×10^{-5}	1.5×10^{-8}
After NN cut ($NN > 0.6$)	5.5%	2.0×10^{-5}	3.0×10^{-8}
After BDT cut ($BDT > 0.5$)	5.7%	2.1×10^{-5}	3.2×10^{-8}

Table 5: Comparison among the efficiencies of the neural network, boosted decision tree methods and hybrid-R reconstruction method by step-by-step cuts for $M_{H^\pm} = 1 \text{ TeV}$.

It is found that the signal and background separation can be improved when the mass of the charged Higgs boson increases. From Fig. 10, it is reliable to conclude that the whole mass range of heavy charged Higgs boson in our study can be feasible at the LHC.

5. Conclusions

In this paper, we have studied the full hadronic decay mode of the heavy charged Higgs production process $pp \rightarrow t_2 H^- \rightarrow t_2 \bar{t}_1 b_3$. The main task to reconstruct the heavy charged Higgs boson in association with a non-boosted top comes from the large combinatorics and the huge QCD background events. By using the b taggings and top tagger methods to suppress the background from the SM, we propose a hybrid-R reconstruction method to fully reconstruct the signal.

Our hybrid-R reconstruction method exploits the advantages of a larger and a smaller cone size of jet algorithms and can greatly reduce the combinatorics. As demonstrated in our analysis, this method is helpful to confirm the signature as well as to separate the signal from the background.

A set of delicate kinematic observables are applied to the multi-variate analysis. Compared with the simple cut method, an improvement around 100% in the S/\sqrt{B} can be achieved. The sensitivity of the charged Higgs boson mass m_{H^\pm} from 800 GeV to 1.5 TeV

	signal $m_{H^\pm} = 1.5 \text{ TeV}$	$t\bar{t} + \text{jets}$	QCD $n_j \geq 4$
After simple cuts	1.8%	1.9×10^{-6}	1.8×10^{-10}
After NN cut ($NN > 0.7$)	4.0%	2.0×10^{-6}	3.6×10^{-10}
After BDT cut ($BDT > 0.5$)	4.5%	2.1×10^{-6}	2.4×10^{-10}

Table 6: Comparison among the efficiencies of the neural network, boosted decision tree methods and hybrid-R reconstruction method by step-by-step cuts for $M_{H^\pm} = 1.5 \text{ TeV}$.

$m_{H^\pm} \text{ (TeV)}$	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5
$\sigma \text{ (fb)}$	324	192	118	75	49	33	22	15
$\frac{S}{\sqrt{B}}$ (with two b taggings & TMVA)	19.4	13.3	9.5	7.0	5.1	4.0	3.1	2.6
lower bound on $\sigma(fb)$	33	29	25	21	19	16	14	11
$\frac{S}{\sqrt{B}}$ (with two b taggings without TMVA)	9.3	6.5	4.6	3.4	2.5	1.9	1.5	1.3
lower bound on $\sigma(fb)$	70	59	51	44	39	35	29	23

Table 7: The significance and sensitivity of LHC for heavy charged Higgs boson production are shown, where we assume the total integrated luminosity as 100 fb^{-1} . For the sake of comparison, the results with and without TMVA are provided. We use $\frac{S}{\sqrt{B}} = 2.0$ to obtain the exclusion lower bound on the cross section.

is also explored and it is found that the LHC can probe the heavy charged Higgs boson mass up to 1.5 TeV. When the integrated luminosity is higher, even heavier mass region of the charged Higgs boson can be reachable.

In this paper, we chose a reference THDM model as an example to perform a concrete study and the cross section of heavy charged Higgs boson is approximately proportional to $Y_L^2 + Y_R^2$. To apply our results to other models, like the supersymmetry model, one can rescale the cross section and the significance in term of the model parameters.

In the evaluation of cross section of signal, we have included the process $gb \rightarrow \bar{b}tH^-$ which is a part of QCD next-to-leading (NLO) order correction of process $gb \rightarrow tH^-$ and we only assume the K-factor as 1.5 here. A lot of works on high order QCD corrections and SUSY corrections for process $bg \rightarrow tH^-$ have been done in the past, including QCD NLO calculations [66, 67], SUSY-QCD NLO calculations [67], Yukawa and SUSY electroweak corrections [68] as well as high-order soft gluon corrections [69, 70, 71]. These calculations show the reduction of scale dependence and a significant enhancement to lowest order result when high order corrections are considered. For $\mu = m_{H^\pm}$, $\tan\beta = 30$ and $m_{H^\pm} = 800 \text{ GeV}$ (1 TeV), the K-factor in QCD next-to-next-to-next-to leading order soft gluon threshold correction at next-to-leading logarithmic accuracy is 1.75 (1.81) [70, 71]. Therefore when our results are applied to other models, the significance can be better.

Obviously, our hybrid-R reconstruction method can be generalized to cases where some jets are decay products of one highly boosted massive particle while some of others come from a massive particle near the threshold. Such cases can exist in some new physics

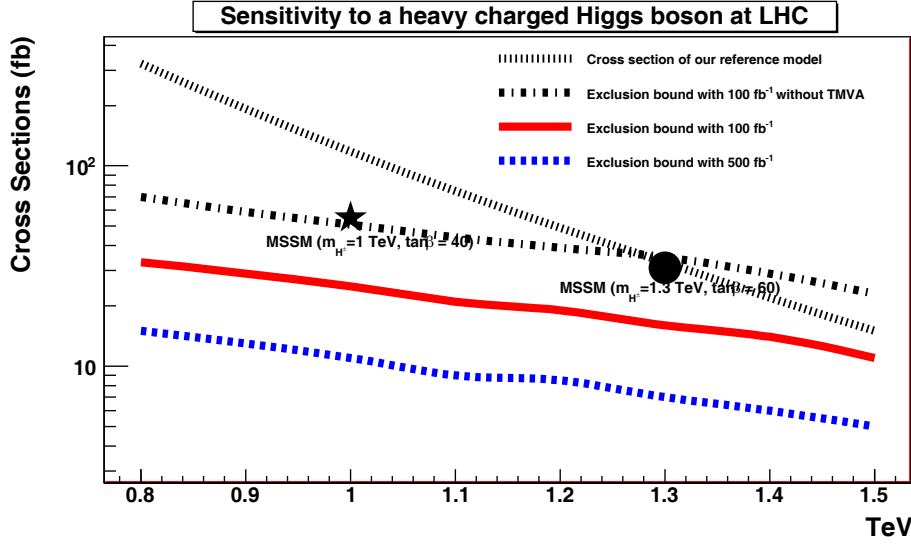


Figure 10: The sensitivity curves of LHC to the heavy charged Higgs bosons with two b tagging are shown from both 100 fb^{-1} and 500 fb^{-1} datasets. The solid star and solid circle markers as two example points of MSSM are also shown.

model where the production of a new heavy particles X in association with a SM massive particle Y , where X decays to boosted massive particles of the SM, such as W bosons, Z bosons, H bosons or top quarks. For examples, the hybrid-R reconstruction method can be applied to the processes $pp \rightarrow W' H \rightarrow tbH$ and $pp \rightarrow t' W \rightarrow WbW$ (where W' and t' represent a new gauge boson and a new up-type quark, respectively). Additionally, our analysis can be extended to the semi-leptonic and di-leptonic modes [72] and to the intermediate heavy charged Higgs boson region (say from 300 GeV to 800 GeV) with the help of HEPToptagger.

Acknowledgments

We would like to thank C.P. Yuan and S. H. Zhu for helpful discussions. This work was supported in part by the National Science Foundation of China under Grant Nos. 11175251. SY would like to thank the hospitality of the “2011 IPMU-YITP School and Workshop on Monte Carlo Tools for LHC”, and “2011 CTEQ Workshop and Conference” during this work. The lectures and discussions on these activities are helpful to this work.

References

- [1] J. F. Gunion, H. E. Haber, G. L. Kane, S. Dawson, *The Higgs Hunter's Guide*, Front. Phys. **80** (2000) 1-448.
- [2] M. Acciarri *et al.*, [L3 Collaboration], *Search for charged Higgs bosons in e^+e^- collisions at center center-of-mass energies up to 202-GeV*, *Phys. Lett. B* **496** (2000) 34 [[hep-ex/0009010](#)];

- ALEPH collaboration, *Search for charged Higgs bosons in e^+e^- collisions at energies up to $\sqrt{S} = 189\text{GeV}$* , *Phys. Lett. B* **487** (2000) 253 [hep-ex/0008005].
- [3] A. Abulencia *et al.* [CDF Collaboration], *Search for charged Higgs bosons from top quark decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\text{-TeV}$* , *Phys. Rev. Lett.* **96** (2006) 042003 [hep-ex/0510065];
V. M. Abazov *et al.* [D0 Collaboration], *Search for charged Higgs bosons in top quark decays*, *Phys. Lett. B* **682** (2009) 278 [arXiv:0908.1811 [hep-ex]];
T. Aaltonen *et al.* [CDF Collaboration], *Search for charged Higgs bosons in decays of top quarks in p anti- p collisions at $\sqrt{s} = 1.96\text{ TeV}$* , *Phys. Rev. Lett.* **103** (2009) 101803 [arXiv:0907.1269 [hep-ex]];
V. M. Abazov *et al.* [D0 Collaboration], *Search for charged Higgs bosons in decays of top quarks*, *Phys. Rev. D* **80** (2009) 051107 [arXiv:0906.5326 [hep-ex]].
- [4] ATLAS Technical proposal, CERN-LHCC-94-38, 254pp. (1994); CMS Technical proposal, CERN-LHCC-94-43, 272pp (1994).
- [5] G. P. Salam, *Towards Jetography*, *Eur. Phys. J. C* **67** (2010) 637 [arXiv:0906.1833 [hep-ph]].
- [6] C. K. Vermilion, *Jet Substructure at the Large Hadron Collider: Harder, Better, Faster, Stronger*, [arXiv:1101.1335 [hep-ph]].
- [7] A. Abdesselam, E. B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, P. Bruckman de Renstrom, D. Buarque Franzosi, R. Buckingham *et al.*, *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur. Phys. J. C* **71** (2011) 1661 [arXiv:1012.5412 [hep-ph]].
- [8] L. G. Almeida, R. Alon and M. Spannowsky, *Structure of Fat Jets at the Tevatron and Beyond*, arXiv:1110.3684 [hep-ph].
- [9] J.F. Gunion, *Detecting the tb Decays of a Charged Higgs Boson at a Hadron Supercollider*, *Phys. Lett. B* **322** (1994) 125 [hep-ph/9312201].
- [10] V. Barger, R.J. Phillips, D.P. Roy, *Heavy Charged Higgs Signals at the LHC*, *Phys. Lett. B* **324** (1994) 236 [hep-ph/9311372].
- [11] S. Moretti, D.P. Roy, *Detecting heavy charged Higgs bosons at the LHC with triple b tagging*, *Phys. Lett. B* **470** (1999) 209 [hep-ph/9909435];
S. Moretti, K. Odagiri, *Production of charged Higgs bosons of the minimal supersymmetric standard model in b quark initiated processes at the large hadron collider*, *Phys. Rev. D* **55** (1997) 5627 [hep-ph/9611374];
C.S. Huang, S.H. Zhu, *Supersymmetric Higgs bosons discovery potential at hadron colliders through $b g$ channel*, *Phys. Rev. D* **60** (1999) 075012 [hep-ph/9812201];
A. Belyaev, D. Garcia, J. Guasch, J. Sola, *Prospects for heavy supersymmetric charged Higgs boson searches at hadron colliders*, *J. High Energy Phys.* **06** (2002) 059 [hep-ph/0203031].
- [12] D.J. Miller, S. Moretti, D.P. Roy, W.J. Stirling, *Detecting heavy charged Higgs bosons at the CERN LHC with four b quark tags*, *Phys. Rev. D* **61** (2000) 055011 [hep-ph/9906230].
- [13] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, *Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks*, *Phys. Rev. Lett.* **101** (2008) 142001 [arXiv:0806.0848 [hep-ph]].
- [14] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Longitudinally invariant $K(t)$ clustering algorithms for hadron hadron collisions*, *Nucl. Phys. B* **406** (1993) 187.

- [15] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev. D* **48** (1993) 3160 [hep-ph/9305266]
- [16] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better Jet Clustering Algorithms*, *J. High Energy Phys.* **08** (1997) 001 [hep-ph/9707323].
- [17] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross sections in deep-inelastic scattering* [hep-ph/9907280].
- [18] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, *J. High Energy Phys.* **04** (2008) 063 [arXiv:0802.1189 [hep-ph]]
- [19] W. Skiba and D. Tucker-Smith, *Using jet mass to discover vector quarks at the LHC*, *Phys. Rev. D* **75** (2007) 115010 [hep-ph/0701247].
- [20] B. Holdom, *The heavy quark search at the LHC*, *J. High Energy Phys.* **08** (2007) 069 [arXiv:0705.1736 [hep-ph]].
- [21] K. Agashe *et al.*, *LHC Signals for Warped Electroweak Neutral Gauge Bosons*, *Phys. Rev. D* **76** (2007) 115015 [arXiv:0709.0007 [hep-ph]].
- [22] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, *New clustering algorithm for multi - jet cross-sections in e^+e^- annihilation*, *Phys. Lett. B* **269** (1991) 432.
- [23] M. H. Seymour, *Searches for new particles using cone and cluster jet algorithms: A Comparative study*, *Z. Phys. C* **62** (1994) 127.
- [24] J. M. Butterworth, B. E. Cox and J. R. Forshaw, *W W scattering at the LHC*, *Phys. Rev. D* **65** (2002) 096014 [hep-ph/0201098].
- [25] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001 [arXiv:0802.2470 [hep-ph]].
- [26] J. Thaler and L. T. Wang, *Strategies to Identify Boosted Tops*, *J. High Energy Phys.* **07** (2008) 092 [arXiv:0806.0023 [hep-ph]].
- [27] M. M. Nojiri and M. Takeuchi, *Study of the top reconstruction in top-partner events at the LHC*, *J. High Energy Phys.* **10** (2008) 025 [arXiv:0802.4142 [hep-ph]].
- [28] T. Plehn, G. P. Salam and M. Spannowsky, *Fat Jets for a Light Higgs*, *Phys. Rev. Lett.* **104** (2010) 111801 [arXiv:0910.5472 [hep-ph]].
- [29] D. Krohn, J. Thaler and L. T. Wang, *Jet Trimming*, *J. High Energy Phys.* **02** (2010) 084 [arXiv:0912.1342 [hep-ph]].
- [30] S. D. Ellis, C. K. Vermilion and J. R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, *Phys. Rev. D* **80** (2009) 051501 [arXiv:0903.5081 [hep-ph]].
- [31] S. D. Ellis, C. K. Vermilion and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev. D* **81** (2010) 094023 [arXiv:0912.0033 [hep-ph]].
- [32] J. M. Butterworth, A. Davison, E. Ozcan, and P. Sherwood, *Ysplitter : An athena tool for studying jet substructure*, ATL-PHYS-INT-2007-015.
- [33] ATLAS Collaboration, *ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta*, ATL-PHYS-PUB-2009-088.

- [34] G. Brooijmans, *High p_T Hadronic Top Quark Identification, Part I: Jet Mass and YSplitter*, ATLAS note, ATL-COM-PHYS-2008-001.
- [35] ATLAS Collaboration, *Reconstruction of High Mass tt Resonances in the Lepton+Jets Channel*, ATL-PHYS-PUB-2009-081.
- [36] CMS Collaboration, *A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging*, CMS-PAS-JME-09-001.
- [37] D. Krohn, J. Thaler, L. -T. Wang, *Jets with Variable R* , *J. High Energy Phys.* **06** (2009) 059 [arXiv:0903.0392 [hep-ph]].
- [38] G. D. Kribs, A. Martin, T. S. Roy and M. Spannowsky, *Discovering the Higgs Boson in New Physics Events using Jet Substructure*, *Phys. Rev. D* **81** (2010) 111501 [arXiv:0912.4731 [hep-ph]].
- [39] G. D. Kribs, A. Martin, T. S. Roy and M. Spannowsky, *Discovering Higgs Bosons of the MSSM using Jet Substructure*, *Phys. Rev. D* **82** (2010) 095012 [arXiv:1006.1656 [hep-ph]].
- [40] J. Thaler, K. Van Tilburg, *Identifying Boosted Objects with N -subjettiness*, *J. High Energy Phys.* **03** (2011) 015 [arXiv:1011.2268 [hep-ph]].
- [41] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, I. Sung and J. Virzi, *Substructure of high- p_T Jets at the LHC*, *Phys. Rev. D* **79** (2009) 074012 [arXiv:0807.0234 [hep-ph]].
- [42] L. G. Almeida, S. J. Lee, G. Perez, I. Sung and J. Virzi, *Top Jets at the LHC*, *Phys. Rev. D* **79** (2009) 074012 [arXiv:0810.0934 [hep-ph]].
- [43] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, *Stop Reconstruction with Tagged Tops*, *J. High Energy Phys.* **10** (2010) 078 [arXiv:1006.2833 [hep-ph]].
- [44] T. Plehn, M. Spannowsky and M. Takeuchi, *Boosted Semileptonic Tops in Stop Decays*, *J. High Energy Phys.* **05** (2011) 135 [arXiv:1102.0557 [hep-ph]].
- [45] G. D. Kribs, A. Martin and T. S. Roy, arXiv:1012.2866 [hep-ph].
- [46] V. Barger, P. Huang, *Hollow cone sieve for top tagging*, [arXiv:1110.2214 [hep-ph]].
- [47] A.A. Barrientos Bendezú, B.A. Kniehl, *$W^\pm H^\mp$ associated production at the large hadron collider*, *Phys. Rev. D* **59** (1999) 015009 [hep-ph/9807480];
S. Moretti, K. Odagiri, *The Phenomenology of $W^\pm H^\mp$ production at the large hadron collider*, *Phys. Rev. D* **59** (1999) 055008 [hep-ph/9809244];
O. Brein, W. Hollik, S. Kanemura, *The MSSM prediction for $W^\pm H^\mp$ production by gluon fusion*, *Phys. Rev. D* **63** (2001) 095001 [hep-ph/0008308];
S. -S. Bao, X. Gong, H. -L. Li, S. -Y. Li and Z. -G. Si, *Identify Charged Higgs Boson in $W^\pm H^\mp$ Associated Production at LHC*, [arXiv:1112.0086 [hep-ph]].
- [48] S. Kanemura, C. P. Yuan, *Testing supersymmetry in the associated production of CP odd and charged Higgs bosons*, *Phys. Lett. B* **530** (2002) 188 [hep-ph/0112165];
Q. -H. Cao, S. Kanemura, C. P. Yuan, *Associated production of CP odd and charged Higgs bosons at hadron colliders*, *Phys. Rev. D* **69** (2004) 075008 [hep-ph/0311083];
E. Asakawa, O. Brein and S. Kanemura, "Enhancement of $W^\pm H^\mp$ production at hadron colliders in the two Higgs doublet model," *Phys. Rev. D* **72** (2005) 055017 [hep-ph/0506249];
D. Eriksson, S. Hesselbach and J. Rathsman, "Associated charged Higgs and W boson production in the MSSM at the CERN Large Hadron Collider," *Eur. Phys. J. C* **53** (2008) 267 [hep-ph/0612198].

- [49] S.S. Willenbrock, *Pair Production Of Supersymmetric Charged Higgs Bosons*, *Phys. Rev. D* **35** (1987) 173;
A. Krause, T. Plehn, M. Spira, P.M. Zerwas, *Production of charged Higgs boson pairs in gluon-gluon collisions*, *Nucl. Phys. B* **519** (1998) 85 [hep-ph/9707430];
A.A. Barrientos Bendezi, B.A. Kniehl, *H^+H^- pair production at the Large Hadron Collider*, *Nucl. Phys. B* **568** (2000) 305 [hep-ph/9908385];
O. Brein, W. Hollik, *Pair production of charged MSSM Higgs bosons by gluon fusion*, *Eur. Phys. J. C* **13** (2000) 175 [hep-ph/9908529].
- [50] J. Conway, “PGS4”, Pretty Good Simulation of high energy collision,
<http://physics.ucdavis.edu/~conway/research/software/pgs/pgs4-general.htm>
- [51] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, *J. High Energy Phys.* **02** (2003) 027 [hep-ph/0208156].
- [52] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *J. High Energy Phys.* **07** (2003) 001. [hep-ph/0206293].
- [53] T. Sjostrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, *J. High Energy Phys.* **05** (2006) 026 [hep-ph/0603175].
- [54] M. Cacciari, G. Salam, and G. Soyez, “FastJet” [<http://fastjet.fr/>].
- [55] Pierre-Antoine Delsart, Kurtis Geerlings, Joey Huston, Brian Martin, and Christopher Vermilion, “SpartyJet” [<http://projects.hepforge.org/spartyjet/>]
- [56] M. Cacciari, G. P. Salam, G. Soyez, *The Catchment Area of Jets*, *J. High Energy Phys.* **04** (2008) 005 [arXiv:0802.1188 [hep-ph]].
- [57] M. A. Pleier, *Review of Top Quark Properties Measurements at the Tevatron*, *Int. J. Mod. Phys. A* **24** (2009) 2899 [arXiv:0810.5226 [hep-ex]].
- [58] T. Aaltonen *et al.* [CDF Collaboration], *Measurement of the $p\bar{p} \rightarrow t\bar{t}$ production cross-section and the t mass at $\sqrt{s} = 1.96\text{-TeV}$ in the all-hadronic decay mode*, *Phys. Rev. D* **76** (2007) 072009 [arXiv:0706.3790 [hep-ex]].
- [59] V. M. Abazov *et al.* [D0 Collaboration], *Measurement of the $p\bar{p} \rightarrow t\bar{t}$ production cross section at $\sqrt{s} = 1.96\text{-TeV}$ in the fully hadronic decay channel*, *Phys. Rev. D* **76** (2007) 072007 [arXiv:hep-ex/0612040].
- [60] Y. Cui, Z. Han, M. D. Schwartz, *W-jet Tagging: Optimizing the Identification of Boosted Hadronically-Decaying W Bosons*, *Phys. Rev. D* **83** (2011) 074023 [arXiv:1012.2077 [hep-ph]].
- [61] J. Gallicchio, J. Huth, M. Kagan, M. D. Schwartz, K. Black, B. Tweedie, *Multivariate discrimination and the Higgs + W/Z search*, *J. High Energy Phys.* **04** (2011) 069 [arXiv:1010.3698 [hep-ph]].
- [62] J. A. Aguilar-Saavedra, *Identifying top partners at LHC*, *J. High Energy Phys.* **11** (2009) 030. [arXiv:0907.3155 [hep-ph]].
- [63] B. Holdom, Q. -S. Yan, *The b' search at the LHC*, [arXiv:1004.3031 [hep-ph]] (to be published in PRD).
- [64] B. Holdom, Q. -S. Yan, *Searches for the t' of a fourth family*, *Phys. Rev. D* **83** (2011) 114031 [arXiv:1101.3844 [hep-ph]].

- [65] A. Ali, F. Barreiro, J. Llorente, *Improved sensitivity to charged Higgs searches in Top quark decays $t \rightarrow bH^+ \rightarrow b(\tau^+\nu_\tau)$ at the LHC using τ polarisation and multivariate techniques*, *Eur. Phys. J. C* **71** (2011) 1737 [arXiv:1103.1827 [hep-ph]].
- [66] S. H. Zhu, *Complete next-to-leading order QCD corrections to charged Higgs boson associated production with top quark at the CERN large hadron collider*, *Phys. Rev. D* **67** (2003) 075006 [hep-ph/0112109];
- [67] T. Plehn, *Charged Higgs boson production in bottom gluon fusion*, *Phys. Rev. D* **67** (2003) 014018 [hep-ph/0206121];
E. L. Berger, T. Han, J. Jiang, T. Plehn, *Associated production of a top quark and a charged Higgs boson*, *Phys. Rev. D* **71** (2005) 115012 [hep-ph/0312286].
- [68] L. G. Jin, C. S. Li, R. J. Oakes, S. H. Zhu, *Yukawa corrections to charged Higgs boson production in association with a top quark at hadron colliders*, *Phys. Rev. D* **62** (2000) 053008 [hep-ph/0003159];
M. Beccaria, G. Macorini, L. Panizzi, F. M. Renard, C. Verzegnassi, *Associated production of charged Higgs and top at LHC: The Role of the complete electroweak supersymmetric contribution*, *Phys. Rev. D* **80** (2009) 053011 [arXiv:0908.1332 [hep-ph]].
- [69] N. Kidonakis, *Next-to-next-to-next-to-leading-order soft-gluon corrections in hard-scattering processes near threshold*, *Phys. Rev. D* **73** (2006) 034001 [hep-ph/0509079].
- [70] N. Kidonakis, *Charged Higgs production via $bg \rightarrow tH^-$ at the LHC*, *J. High Energy Phys.* **05** (2005) 011 [hep-ph/0412422].
- [71] N. Kidonakis, *Higher order corrections to H^\pm production* PoS **CHARGED2008** (2008) 003 [arXiv:0811.4757 [hep-ph]].
- [72] Q. -S Yan, S. Yang, in preparation.